

DOCUMENT RESUME

ED 334 655

EA 023 121

TITLE Teacher Evaluation Programs in SREB States.
INSTITUTION Southern Regional Education Board, Atlanta, Ga.
PUB DATE 90
NOTE 23p.
PUB TYPE Reports - Descriptive (141) -- Reports -
Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Educational Change;
Educational Policy; Elementary Secondary Education;
*Evaluation Criteria; *Program Evaluation; Skills;
*State Action; *Teacher Evaluation
IDENTIFIERS *United States (South)

ABSTRACT

Working with state education agency staff and higher education faculty, the Southern Regional Education Board (SREB) attempted to assemble information about current state-developed teacher evaluation systems and future directions for teacher assessment. SREB states were trying to determine whether: (1) the same evaluation systems were being used from state to state; (2) the SREB states had a common evaluation language; (3) decisions about beginning and veteran teachers differed; and (4) states' conclusions about good teaching varied. Of the 12 southern states participating in the study, 11 have developed evaluation programs for both beginning and veteran teachers. The study was designed in two phases. In phase 1, researchers examined documents describing each state's evaluation program and prepared a written analysis. In phase 2, trained observation teams from each state used their own evaluation systems to evaluate the same set of videotapes of classroom teaching. The SREB study team's comparative analysis yielded the following conclusions: (1) states are developing some common understanding about observable teacher behaviors and a common language for observing them; (2) a basis exists for translating observation information gathered from state to state; and (3) all states could benefit from additional research aimed at improving observation systems. The "next steps" and implications for state policy are also summarized. (MLH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED834655

2

Teacher Evaluation Programs in SREB States

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.
• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
M. A. Sullivan

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

EA 023 121

TEACHER EVALUATION PROGRAMS
IN SREB STATES

Southern Regional Education Board
592 Tenth Street, NW
Atlanta, Georgia 30318-5790
(404) 875-9211

1990

FOREWORD

This publication reports on the results of a joint effort by twelve SREB states to examine and seek ways to improve their teacher evaluation systems. It also provides a textbook example of the value of cooperative ventures among states. During the course of the project, higher education faculty, state department of education staff, teachers, principals, and the Southern Regional Education Board all contributed resources and expertise.

The leadership provided by the University of Tennessee and its College of Education, and by the North Carolina Department of Public Instruction, has been exemplary and illustrates what can be accomplished when state leaders are willing to take an initiative that has implications beyond their own states' borders.

This report offers important information about how teachers are currently evaluated and about the strengths and weaknesses of evaluation programs now in place. States are under mounting pressure to stretch tax dollars and find the most cost-effective ways possible to maintain educational quality. It makes sense for states to improve teacher evaluation programs by drawing on the experience and knowledge in other states where similar work is underway. States should also consider the benefits of linking programs regionally to eliminate the needless re-evaluation of experienced teachers moving from state to state. Not only will states save money and time by accepting similar evaluations from other states, they will remove another barrier that discourages experienced teachers from maintaining their certification when they cross state borders.

Mark D. Musick
President

INTRODUCTION

Southern Regional Education Board (SREB) states led the educational reform efforts of the 1980s with comprehensive state programs to improve education. Many reform packages included initiatives to evaluate teacher performance. These evaluations were designed in part to answer questions about the quality of the teacher work force.

Legislatures appropriated funds and left the task of developing and implementing teacher evaluation programs to state and local education agencies. A flurry of simultaneous, similar, often parallel activities were undertaken from state to state and from school district to school district. As a result, a multitude of teacher evaluation systems, often with different purposes, were developed and implemented across the SREB region.

The extensive work in teacher performance evaluation during the 1980s reflected attempts to resolve in new ways a long-running debate about teachers that centered on three issues: teacher accountability versus teacher assistance; teacher performance versus student achievement; and individual teacher growth versus the organizational needs of a school or school district.

The issues surrounding the evaluation of teacher performance are not confined to individual states. They span the region and the nation. How do we find out more about the link between what teachers do in the classroom and how students learn? How do we refine teacher evaluation systems? How can we help states include the results of new-teacher evaluations in certification reciprocity agreements and eliminate the need to re-evaluate teachers as they move from state to state?

SREB, working with state education agency staff and higher education faculty, set out to assemble information about current state-developed teacher evaluation systems and future directions for teacher assessment. SREB states were seeking answers to key questions:

- Are the same evaluation criteria being used from state to state?
- Is there a common language of teacher evaluation in the SREB states?

- Do states using different evaluation systems make similar decisions about whether beginning teachers demonstrate teaching skills for regular certification?

- Are the decisions made about veteran teachers (e.g., for continuing employment or incentive pay) the same from state to state?

- Do different evaluation systems reach similar conclusions about what good teaching is?

- Are there ways for states to work together to improve the evaluation of teacher performance?

The study was directed by Dr. Russell French of the University of Tennessee in Knoxville in cooperation with state department of education personnel, with additional staff work by Dr. David Holdzkorn and Dr. Barbara Kuligowski of the North Carolina Department of Public Instruction.

Twelve states participated in one or both phases of the study. Of those 12 states, one (Virginia) has developed or mandated an evaluation program only for beginning teachers (1-3 years of experience). The other 11 states (Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, Oklahoma, Tennessee, Texas, and West Virginia) report programs for beginning *and* experienced teachers.

The study was designed in two phases. Phase I was carried out by examining documents that described each state's evaluation program in detail and preparing a written analysis, which each state reviewed for accuracy.

In Phase II, trained observation teams from each state used their own evaluation systems to evaluate the same set of videotapes of classroom teaching. The SREB study team compared and analyzed the decisions made by observers using various state evaluation programs and drew some tentative conclusions about comparability.

Part 1
HOW SIMILAR AND HOW DIFFERENT ARE
TEACHER EVALUATION SYSTEMS IN SREB
STATES?

Purposes

❑ Teacher evaluation systems in SREB states have been designed to serve two groups of teachers (beginning and experienced) and have at least five different purposes. The most common purpose for *beginning* teacher evaluation in SREB states is certification. The most common purpose for *continuing* teacher evaluation in SREB states is instructional improvement.

Sources

❑ SREB states have drawn upon common sources and used many of the same processes in establishing teacher evaluation criteria. The three most prevalent sources of criteria have been effective teaching research, consensus of teachers, and job analyses.

❑ There is little evidence that emerging research in instruction (e.g., results of inductive methods, group processes) has yet become a part of teacher performance criteria.

❑ The criteria SREB states have developed consistently reflect sensitivity to the issues of teacher involvement and legal defensibility — sensitivity that was not present in teacher evaluation a decade ago.

❑ While criticism is sometimes leveled at current evaluation systems for focusing on teacher behaviors related to "direct teaching," these systems reflect the facts that (a) research findings are legally defensible, while theory is not; and (b) educators readily agree upon the value of certain teacher behaviors and practices.

Criteria

☐ Criteria used to judge teacher performance are very similar. Most assess planning, delivery of instruction, evaluation of student progress, classroom management, student involvement, basic communication skills, classroom climate, and interpersonal skills.

☐ There is substantially less agreement about how teaching behaviors are defined or how they are grouped under each criterion.

☐ States that evaluate both beginning and experienced teachers generally use the same assessment criteria. However, data may be gathered in different ways. (For example, one observation may be used in evaluating a continuing teacher, but three observations may be required in the state's beginning teacher evaluation. Or, an interview process may be used to collect information about planning from continuing teachers, while beginning teachers submit lesson plans for review.) Few evaluation programs use different criteria or weight criteria differently for the two groups.

☐ Only five evaluation programs in four states report assessment of innovative teacher practices like cooperative learning. (State representatives report that their evaluation programs try not to inhibit innovative practices, but they do not reward such practices.)

☐ Only nine programs in five states attempt to directly relate student outcomes to teacher evaluation.

☐ Teacher practices that relate to school effectiveness (sharing ideas and materials, initiating activities and projects, assisting peers) are included in teacher evaluation in five states.

Development of Evaluation Systems

☐ Although extensive work in teacher assessment has been going on in the SREB states for more than a decade, a majority of these states have implemented their current programs since 1985.

☐ Teacher evaluation programs have been legislated into existence in 11 of 12 SREB states participating in the teacher evaluation study. In eight states, State Board of Education policies have supported and clarified that legislation.

☐ In developing teacher evaluation criteria, no state used fewer

than five sources. Four different patterns emerged in the development of teacher evaluation in the SREB states:

- A state-developed evaluation system with local implementation;
- State-developed evaluation criteria, with locally developed instruments and procedures, and local implementation;
- A state-developed evaluation system and state implementation;
- A locally developed evaluation system and local implementation, with state assistance.

☐ The local evaluation system developed under state guidelines may represent the trend of the future, with states requiring that districts apply the knowledge of instruction and evaluation now available.

☐ Third party (external) reviews of the current evaluation systems are needed. Only three states have conducted such studies; two other states have them in process.

Observation Procedures

☐ Classroom observation is an important teacher evaluation methodology in all 12 SREB states participating in the teacher evaluation study.

☐ In most evaluation systems studied, the classroom observation generated records of a teacher's actions through a "script" or coding scheme that compared the actions to pre-selected behaviors.

☐ Observation procedures reflect sensitivity to the procedural questions most often raised in evaluation appeals or legal challenges. These questions include: adequacy of documentation, number of observations, length of observations, communication with the person being evaluated, and consistency of procedures across candidates and evaluators.

☐ The observation procedures used in these evaluation systems constitute a dramatic change from many pre-1980 evaluation programs in which teachers were rarely observed, and little attention was given to sound principles of measurement and evaluation.

Observers and Evaluators

☐ While earlier teacher evaluation relied solely on principals as evaluators, there is a trend in SREB states toward the use of multiple observers/evaluators.

☐ Seven of the states participating in the study include teachers in their evaluator teams. This procedure constitutes a significant change from the historical model which designated school administrators as the only teacher evaluators.

☐ Only half of the states have established performance standards for evaluators.

Evaluator Training

☐ Training of observers and evaluators is required in practically all of the state-level teacher evaluation programs.

☐ State Departments of Education currently are the primary developers and providers of evaluator training programs, sometimes called "turnkey" training packages. In most (but not all) cases, training time for evaluators appears to be consistent with the demands of the evaluation system.

☐ There is increasing emphasis on follow-up training for evaluators, probably in recognition of the problem of "evaluator drift" — a tendency of all evaluators to drift away from original definitions over time. Re-training also meets a perceived need to clarify and refine evaluation practices.

Evaluation Procedures Other Than Observation

☐ There is heavy reliance on classroom observation as a source of evaluation data. In three states it is the only information used.

☐ There is a trend in the states toward the use of multiple forms of data (interviews, self reports, administrative records) to assess teachers; nine of 12 states report the use of more than one kind.

☐ Instruments and data collection procedures most often used in addition to observation are candidate interviews, review of administrator records, and candidate self-reports.

Communication With Teachers Who Are Being Assessed

❑ The SREB states participating in this study have invested heavily in communication with teachers who are being assessed. Orientation programs and regular feedback occur in all teacher evaluation models.

❑ About half the state programs attempt to bridge the gap between how teachers are trained and how teachers are evaluated through some form of professional development planning.

What Evaluation Systems Do Well

Teacher evaluation systems in the 12 SREB states participating in this study do a number of things well:

✓ *They use effective teaching research to assess the teaching process.*

All the evaluation programs studied draw heavily upon the effective teaching knowledge base developed in the 1970s and 1980s. Programs focus on best teaching practices as defined in the effective teaching research of that time. New knowledge about teaching is still in the "theory" stage. Research to translate this theory into practical systems of evaluation needs to be funded.

✓ *They invest in beginning teachers.*

Ten states in the study conduct statewide assessments of beginning teachers. In most cases, standards for performance are incorporated into licensure requirements; as a result, most beginning teachers must demonstrate satisfactory performance prior to licensure. In addition to the assessment programs, many SREB states provide assistance to beginning teachers in the form of mentors or coaches. Evaluation data provide the basis for this assistance, linking assessment and induction into the profession.

✓ *They demonstrate consistency in evaluation practices.*

The criteria used to judge teacher performance in the 12 states studied are very similar at one level. Most of the evaluation programs assess teacher planning, delivery of instruction, teacher evaluation of student progress, classroom management, student involvement in the teaching/learning process, teacher communication skills, classroom climate, and teacher interpersonal skills. When assigning specific teacher behaviors to these competency areas, there is somewhat less

agreement, and there is not total agreement on the definitions of behaviors and practices.

In addition to the consistency found in criteria specification, there is also great consistency found in the development and implementation procedures used to ensure fairness, objectivity and legal defensibility.

✓ *They establish a commonality of language and of the concepts of teaching.*

Individuals who participated in this study had little difficulty in understanding the questions posed and the terminology used by investigators. Nor did they find it difficult to cluster criteria and procedures as requested. In addition, state representatives report that a common language and conceptualization of teaching has developed within their states. Obviously, there is some uniformity of language and concept both within and across states.

✓ *They establish new forms of professional development.*

Evaluation orientation programs, pre- and post-observation conferences, and evaluator training programs also serve as professional development programs. Thousands of teachers and administrators in these 12 states have now participated in these programs. Many participants indicated that they did not really understand instruction until they learned how to evaluate it and discuss it with others.

✓ *They establish new links between evaluation and professional development.*

While the potential for linking evaluation results with professional development programs has always existed, that linkage has seldom been established. Many of the programs have established the linkage by asking that individual professional development plans be developed. There is now more concern about the delivery of formal staff development programs and activities that will address weaknesses found among groups of teachers. For instance, if classroom management is a weakness that is revealed by evaluation of beginning teachers, then state training can focus on that knowledge and skill.

What Evaluation Systems Do Not Do Well

While the time and resources given to teacher evaluation in the SREB states over the past decade have accomplished much, this study suggests that there are areas in which the current evaluation systems may be improved. Here are some areas of concern:

✖ The programs do not assess the teacher's knowledge of content well.

National discussion is underway about the teacher's knowledge of content and his or her ability to apply that knowledge to the range of learners and situations a teacher is likely to encounter. While most of the evaluation systems analyzed in this study address "teacher coverage of content" in some way, the most common tool for assessment is classroom observation by an observer who is not a specialist in the content being taught. If a state or school district desires to know the teacher's knowledge of content and his/her ability to teach content appropriately to a range of learners, the observers may not be able to assess it. Observation may not be the best assessment tool for this purpose.

✖ The programs generally do not assess the relationship between teacher practices and student outcomes.

Only a few of the participating states attempt to assess changes in student performance (achievement, attitudes, motivation, etc.) and link these to teacher performance. The argument most often used in programs that do not attempt to assess student achievement is that documenting the teacher's use of behaviors that are known to correlate with student achievement is as close as the evaluation process can or should come. However, the question remains whether that argument will satisfy the general public and state policymakers who called for teacher evaluations.

✖ The programs do not systematically document teacher performance in areas that are not observable in the classroom.

Only three of the evaluation systems studied rely solely on classroom observation for data collection/performance documentation. The use of assessment techniques other than classroom observation is erratic, despite the presence of criteria in most programs that clearly require data from different sources.

✖ The programs do not distinguish between good and best teaching.

Little attention appears to be given to determining the *quality* of instruction, except in three states that are implementing career ladder evaluation programs. This finding is further supported by the lack of attention given to assessments that would include higher level expectations for experienced teachers.

✱ *The programs are not evaluated systematically.*

Only a few states have subjected their evaluation systems to third-party evaluations. The procedures used for establishing the validity, reliability, credibility, and impact of the systems vary greatly from state to state. Insufficient attention has been given to developing ways to evaluate the systems and their impact. In fairness, it should be noted that most of these evaluation programs have been implemented within the last five years, and early efforts and resources had to be focused on system development and implementation.

Part II: CLASSROOM OBSERVATION OF TEACHERS

If classroom observations are made of a teacher in *State A*, using that state's evaluation procedures, will you come to the same conclusions if you use *State B*'s observation procedures? Will the answer to this question depend on whether we are talking about a beginning teacher or a continuing teacher?

Observations of teachers at work in classroom settings are a part of all the evaluation systems reviewed in this study. Education policy makers and researchers will need answers to questions about the equivalency or compatibility of SREB states' observation and evaluation systems as they deal with issues such as increased teacher mobility and improved preparation of teachers.

SREB and its partners in this study began to explore these issues by asking teams of observers in participating states to review and evaluate a set of videotapes of teachers in their classrooms. Time and cost limitations did not permit a definitive comparison. From the outset, the project was intended as an exploratory study *only*. It was undertaken to develop a general understanding of how comparable the various state evaluation systems might be, and whether further research into comparability might be worthwhile.

Analysis Procedures

Teams of observers from 10 of the 12 states that participated in Phase I (Alabama, Arkansas, Florida, Louisiana, Mississippi, North Carolina, Oklahoma, Tennessee, Texas and Virginia) also took part in the second phase of the evaluation project. Each team was constituted as defined by that state's evaluation system. Teams were asked to observe videotapes and make a series of personnel decisions based on their evaluations.

Six videotapes of teachers in classroom situations were viewed by each team. Three of the tapes contained single lessons taught by three different teachers. The other three tapes contained three separate lessons taught by the same teacher. Even in this small group of tapes, an effort was made to represent a mix of grade levels and subject areas (e.g., elementary and secondary, drafting and English). Tapes were not selected unless they represented at least minimal teacher competence. Two of the teachers in the tapes were female; two were male.

Each observer team used the observation instruments and procedures used in its state. In addition, to permit comparison of results across the states, each state observer team completed a form — referred to simply as the *SREB Decision Form*. On this form, the team members rated the teacher in the tape on eight teaching competencies.

Each observer completed this rating process twice: once as if the teacher in the tape were a beginning teacher; and again, as if the teacher were an experienced teacher. In states where the evaluation system is used for only one of these categories, the state team did only the rating for that category. Three states use consensus ratings by observers (Alabama, North Carolina, and Tennessee), and those states provided consensus ratings where appropriate.

Once teams completed their ratings, they were asked to produce a series of personnel decisions or recommendations. For example, if they were rating the teacher in the tape as a beginning teacher, observers were asked (1) if they could recommend continuing employment, and (2) if they would recommend certification. "Personnel recommendations" were gathered on continuing employment, recertification, and career ladder placement.

All of the methods, procedures, and definitions were incorporated in a detailed *Observer Manual*, but no formal training was given to the observers. Each observer was well trained in the evaluation procedure used in his or her state.

Study Limitations

The SREB study of teacher evaluation programs is limited in several ways. First, few states rely completely on observational data in making personnel decisions, although the study considers observational data only. Most state systems base judgments on multiple observations, not on single observations. Time constraints limited observers to a single viewing of each videotape. Resources were not available to train observers in the use of the *SREB Decision Form*. Only four teachers presenting six lessons were observed, and the teaching performances were limited to those which project staff felt reflected "minimal" competence — consequently, the range of teaching behaviors presented was restricted. (Good tapes that had not already been used by states were hard to find.)

Findings

Agreement Among Observers

There was substantial agreement among the observers from each state. Put differently, the observers from each state indicated they saw the same things in the videotapes and rated them similarly.

Also, the observers from one state showed a high degree of similarity in what they rated and how they rated it with observers from the other states. This is no doubt the result of the states drawing on the same research base for the development of their evaluation systems.

Exceptions to this general pattern occurred in some specific instances. In some cases a particular state's system does not provide for rating a particular characteristic, so comparisons with others were not possible. One state had a relatively new system, and its training procedures were still in development. This state's observers showed more differences in judgment among themselves and with observers from other states.

Not surprisingly, there were more differences when a characteristic was not directly observable. For example, if the observers are asked to evaluate "lesson planning," they must infer from the content of the lesson how well the teacher planned — a more subjective process.

There are similarities, as well as differences, in what the state evaluation systems pay attention to and what observers in each state are trained to look for. States also differ in the major intent of their evaluation systems. The study found that, in general, states which share the same notions regarding evaluation theory or the purpose of evaluations tend to produce similar results.

However, when the judgments requested of the observers focus on broad areas of competence or recommendations such as certifying or rehiring a teacher, there is substantial agreement across the states.

Very limited information was obtained in this study on such important issues as the value of repeated observations or how the work experience of observers influences their observations. Another study might focus on these considerations.

Personnel Decisions

In making personnel decisions, such as recommending continuing employment or recertification, there was substantial agreement from one state to another. This suggests strongly that one state could have confidence in accepting the recommendations from another state about the general level of performance of teachers. Specifically, states

participating in this study that wish to use observation data from other participating states as part of their process for making employment or certification decisions could be reasonably comfortable in doing so.

Where differences did occur, they may have been due in part to a lack of training on the SREB form. Some observers may also have had difficulty making decisions about career ladder placement if they did not have such a system in their states. These features could be improved in the future.

Most states appeared to have addressed the potential problem of "subjectivity" which is often raised by individuals unfamiliar with current observation practices. The data collected were generally consistent with identified performance criteria, and ratings given were consistent with the data available.

However, one cannot conclude that the observations of all teachers within a state are consistent. A high level of consistency depends on the quality of training provided for observers and on the procedures used in selecting observers and carrying out the observations.

Career Ladder Levels

There was substantially less agreement among observers and among states when it came to recommending a particular career ladder level for a teacher (Level I, Level II, or Level III), based on observations of the videotapes.

Part of the difficulty arises from the fact that only three of the participating states have experience making career ladder decisions. Each state's career ladder program is different and may define levels differently. Also, it is very likely that there was insufficient information provided about the use of the SREB form. Observers may have been uncertain about how levels are defined, the differences among levels, and the skills necessary to be classified at a particular rung on the career ladder.

Levels of Teaching Quality

The problems observers encountered in making career ladder decisions may reflect a larger problem with current observational systems. Generally, they do not distinguish well between levels of performance above the minimum. This problem may stem from vague or poorly understood definitions of degrees of teaching quality.

Observer teams were in substantial agreement as to who was the most highly skilled teacher, but there was much less agreement about the relative quality of the three other teachers included in the videotaped lessons. This could be the result of teaching samples that reflected little difference among teachers in terms of overall quality.

Or, the method used in the study to define overall quality could be insensitive to subtle differences.

A more detailed and improved study using specially developed videotapes that reflect important qualitative differences would be more revealing, although the question remains as to whether current observation systems provide the tools observers need to make these distinctions. Evaluation systems that use multiple sources of information will probably be needed.

Beginning and Experienced Teachers

Observers made few distinctions based on whether they were asked to rate a particular videotaped episode as a performance by a beginning teacher or an experienced teacher. There was no systematic tendency for observers to rate beginners more or less leniently than experienced teachers, or to have significantly higher expectations for a teacher described as "experienced."

It may be that observers concentrate on describing the teacher first, and later evaluate what they have seen according to the experience of the teacher. Or, it may be that expectations of what experienced teachers should do, do better, or do more often are not clear or of great magnitude. Also, it could be that the tapes used, and the way they were presented, masked some differences. In any case, this is an area that clearly needs additional study.

Conclusions

Each of the SREB states participating in Phase II of this study has its own system for carrying out classroom observations of teachers. Each state's approach differs in philosophy, purpose, and procedures. Yet there are a number of common threads which cross the state lines.

- States are developing some common understanding about teacher behaviors that can be observed and a common language that describes what has been observed.
- A basis exists for translating observation information gathered in one state to another state. By accomplishing such "translations," states could transfer teacher evaluations from one state to another with little or no loss of quality.
- All states could benefit from additional research aimed at improving observation systems within each state. SREB's

limited but broad-based study demonstrates that interstate cooperation could produce system refinements at a relatively low cost, in comparison to the price a single state might have to pay to accomplish the same goal. (States might, for example, collaborate to produce a set of videotapes of teaching that cover the full range of quality, with multiple observations of the same teacher over time.)

Policymakers should be much encouraged by the results of this exploratory study, which suggests that many states now have teacher observation systems that recognize the same basic teacher competencies as systems in other states in the region.

States have been breaking new ground in teacher evaluation, and it should be encouraging to those persons in each state who have developed the evaluation — and to state legislators and board members who initiated or funded this work — that there is strong agreement among states on what to look for in evaluating good teaching.

Policymakers should also recognize that all existing observation systems can and should be made better. Through cooperative efforts of the states, improved systems can be developed faster, more wisely, and at a more modest cost, with confidence that the teacher's performance is being fairly and accurately described.

Part III: NEXT STEPS

What does the study mean for state policy?

This joint effort of higher education, state department of education personnel, teachers and principals in the schools, and the Southern Regional Education Board shows clearly that states are interested in taking bold steps to examine and take action to improve teacher evaluation in the SREB region. *Action* is the key word. Educators and researchers have been willing to use their expertise and resources to look ahead to improve teacher evaluation in the states. While this study has been exploratory, some conclusions seem justifiable:

- During the 1980s, the SREB states developed state-level teacher evaluation programs to replace those that were not adequately based on research and were not legally defensible.
- Some states developed statewide systems, others developed state guidelines for local implementation, but all state programs have a common understanding of teaching and use similar words and concepts to describe teaching.
- Observing the teacher in the classroom, on the job, is the primary method used to evaluate teaching in the SREB states.
- The decisions reached using different state evaluation systems to determine competency for certification are generally comparable, especially for beginning teachers.
- Classrooms have become more open because principals and teachers are involved in teacher evaluation. Decisions are

often made based on the consensus of both teachers and administrators.

- Staff development is now more often linked to the strengths and weaknesses of teachers than it was before the development of statewide evaluation systems in the 1980s.
- Teacher evaluation in the SREB states today primarily focuses on teacher performance, not student achievement.
- SREB states need to improve methods of evaluating the content knowledge of teachers; they need to ensure that evaluation systems are based on the best research; and they need to develop a means to distinguish good teaching from *excellent* teaching.

What are the next steps?

During the 1980s, each state developed its own teacher evaluation system, but the states relied on many common sources of information, research and experience. Federally funded projects provide most of the research used to develop current systems. As a result, today's teacher evaluation programs in the SREB states are more alike than different. This common ground provides an opportunity for states to build on their extensive knowledge of classroom observation and work jointly to improve their systems. The end result need not be a single system of teacher evaluation. But, a close look at the work of the 1980s argues strongly for joint cooperative efforts among the states in the 1990s to share expertise, to save time and money, and to increase options for reciprocity.

State department of education staff members who participated in the project identified several important policy concerns for their states, including the need to improve classroom observation procedures and follow-up; the need to address technical considerations in evaluating teachers; and the need to link teacher evaluation to staff development and incentive programs. Higher education institutions could also play a key role in research (federal efforts have diminished), evaluating teachers, training evaluators, and designing staff development. The relationship between teacher evaluation and teacher certification for beginning and veteran teachers continues to be a major policy issue.

The following are proposed for consideration:

1. Because evaluation decisions for beginning teachers are generally

comparable, the SREB states should explore reciprocity for initial certification that includes performance evaluation.

2. SREB states have a wealth of resources and expertise that have been devoted to the development of teacher evaluation systems in each state. Higher education institutions, state policymakers, and local district personnel should look for ways to share knowledge, experience, and financial resources in a concerted effort to improve evaluation systems through interstate cooperation.
3. Good selection and training are critical for the persons who do the observation of teachers in the classroom. Superior training materials could be developed at a lower cost through cooperative efforts among states.
4. Two concerns seem paramount:
 - Resources should now be concentrated on evaluation systems that distinguish between merely competent teaching and excellent teaching.
 - Statewide evaluation systems must search for ways to include an assessment of student achievement in the evaluation of teacher performance.
5. Higher education could contribute significantly to the further development of teacher evaluation systems without a major investment of new funds by contributing research time of expert faculty, as the University of Tennessee has done in this study. With higher education involvement, states would be in a better position to link teacher evaluation, teacher education, and certification and professional development.